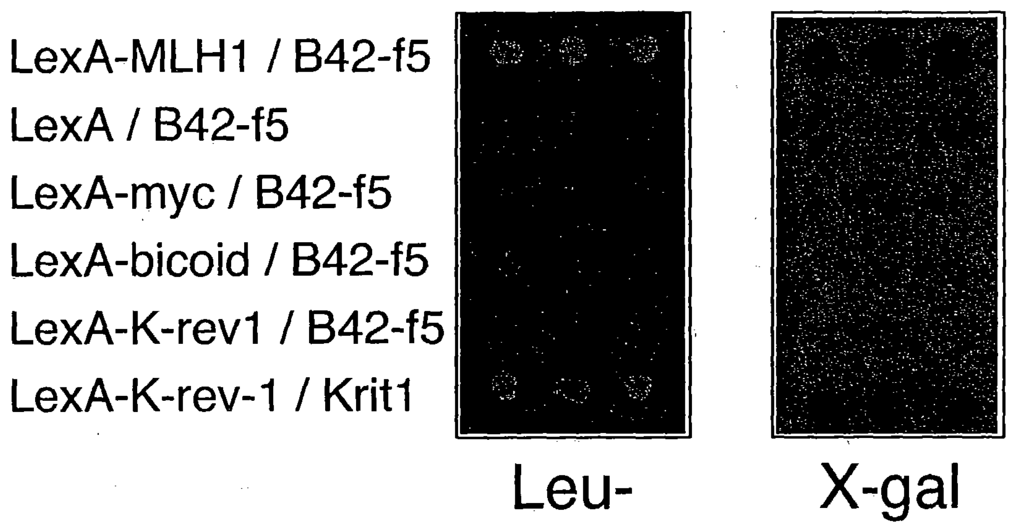

LexA-MLH1 / B42-f5
LexA / B42-f5
LexA-myc / B42-f5
LexA-bicoid / B42-f5
LexA-K-rev1 / B42-f5
LexA-K-rev-1 / Krit1
Leu-
X-gal

Fig. 1

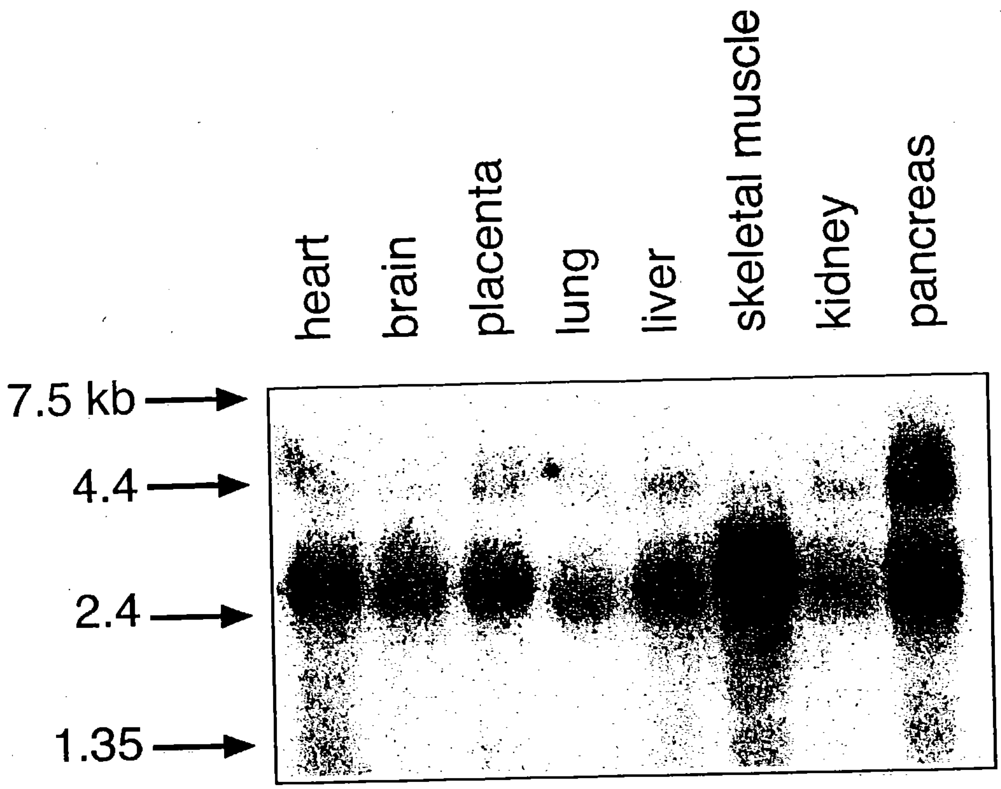

heart
brain
placenta
lung
liver
skeletal muscle
kidney
pancreas
7.5 kb
4.4
2.4
1.35

Fig. 2

```
   1 GGCGGCGTCTGGGGCGCTTTCGCAACATTCAGACCTCGGTTGCAGCCCGGTGCCGTGAGCTGAAGAGGTTTCACATCTTACTCCGCCCCA   90
  91 CACCCTGGGCGTTGCGGCGCTGGGCTCGTTGCTGCAGCCGGACCCTGCTCGATGGGCACGACTGGGCTGGAGAGTCTGAGTCTGGGGGAC  180
                                                                  M  G  T  T  G  L  E  S  L  S  L  G  D

 181 CGCGGAGCTGCCCCCACCGTCACCTCTAGTGAGCGCCTAGTCCCAGACCCGCCGAATGACCTCCGCAAAGAAGATGTTGCTATGGAATTG  270
      R  G  A  A  P  T  V  T  S  S  E  R  L  V  P  D  P  P  N  D  L  R  K  E  D  V  A  M  E  L

 271 GAAAGAGTGGGAGAAGATGAGGAACAAATGATGATAAAAAGAAGCAGTGAATGTAATCCCTTGCTACAAGAACCCATCGCTTCTGCTCAG  360
      E  R  V  G  E  D  E  E  Q  M  M  I  K  R  S  S  E  C  N  P  L  L  Q  E  P  I  A  S  A  Q

 361 TTTGGTGCTACTGCAGGAACAGAATGCCGTAAGTCTGTCCCATGTGGATGGGAAAGAGTTGTGAAGCAAAGGTTATTTGGGAAGACAGCA  450
      F  G  A  T  A  G  T  E  C  R  K  S  V  P  C  G  W  E  R  V  V  K  Q  R  L  F  G  K  T  A

 451 GGAAGATTTGATGTGTACTTTATCAGCCCACAAGGACTGAAGTTCAGATCCAAAAGTTCACTTGCTAATTATCTTCACAAAAATGGAGAG  540
      G  R  F  D  V  Y  F  I  S  P  Q  G  L  K  F  R  S  K  S  S  L  A  N  Y  L  H  K  N  G  E

 541 ACTTCTCTTAAGCCAGAAGATTTTGATTTTACTGTACTTTCTAAAAGGGGTATCAAGTCAAGATATAAAGACTGCAGCATGGCAGCCCTG  630
      T  S  L  K  P  E  D  F  D  F  T  V  L  S  K  R  G  I  K  S  R  Y  K  D  C  S  M  A  A  L

 631 ACATCCCATCTACAAAACCAAAGTAACAATTCAAACTGGAACCTCAGGACCCGAAGCAAGTGCAAAAAGGATGTGTTTATGCCGCCAAGT  720
      T  S  H  L  Q  N  Q  S  N  N  S  N  W  N  L  R  T  R  S  K  C  K  K  D  V  F  M  P  P  S

 721 AGTAGTTCAGAGTTGCAGGAGAGCAGAGGACTCTCTAACTTTACTTCCACTCATTTGCTTTTGAAAGAAGATGAGGGTGTTGATGATGTT  810
      S  S  S  E  L  Q  E  S  R  G  L  S  N  F  T  S  T  H  L  L  L  K  E  D  E  G  V  D  D  V

 811 AACTTCAGAAAGGTTAGAAAGCCCAAAGGAAAGGTGACTATTTTGAAAGGAATCCCAATTAAGAAAACTAAAAAAGGATGTAGGAAGAGC  900
      N  F  R  K  V  R  K  P  K  G  K  V  T  I  L  K  G  I  P  I  K  K  T  K  K  G  C  R  K  S

 901 TGTTCAGGTTTTGTTCAAAGTGATAGCAAAAGAGAATCTGTGTGTAATAAAGCAGATGCTGAAAGTGAACCTGTTGCACAAAAAAGTCAG  990
      C  S  G  F  V  Q  S  D  S  K  R  E  S  V  C  N  K  A  D  A  E  S  E  P  V  A  Q  K  S  Q

 991 CTTGATAGAACTGTCTGCATTTCTGATGCTGGAGCATGTGGTGAGACCCTCAGTGTGACCAGTGAAGAAAACAGCCTTGTAAAAAAAAAA 1080
      L  D  R  T  V  C  I  S  D  A  G  A  C  G  E  T  L  S  V  T  S  E  E  N  S  L  V  K  K  K

1081 GAAAGATCATTGAGTTCAGGATCAAATTTTTGTTCTGAACAAAAAACTTCTGGCATCATAAACAAATTTTGTTCAGCCAAAGACTCAGAA 1170
      E  R  S  L  S  S  G  S  N  F  C  S  E  Q  K  T  S  G  I  I  N  K  F  C  S  A  K  D  S  E

1171 CACAACGAGAAGTATGAGGATACCTTTTTAGAATCTGAAGAAATCGGAACAAAAGTAGAAGTTGTGGAAAGGAAAGAACATTTGCATACT 1260
      H  N  E  K  Y  E  D  T  F  L  E  S  E  E  I  G  T  K  V  E  V  V  E  R  K  E  H  L  H  T

1261 GACATTTTAAAACGTGGCTCTGAAATGGACAACAACTGCTCACCAACCAGGAAAGACTTCACTGGTGAGAAAATATTTCAAGAAGATACC 1350
      D  I  L  K  R  G  S  E  M  D  N  N  C  S  P  T  R  K  D  F  T  G  E  K  I  F  Q  E  D  T

1351 ATCCCACGAACACAGATAGAAAGAAGGAAAACAAGCCTGTATTTTTCCAGCAAATATAACAAAGAAGCTCTTAGCCCCCCACGACGTAAA 1440
      I  P  R  T  Q  I  E  R  R  K  T  S  L  Y  F  S  S  K  Y  N  K  E  A  L  S  P  P  R  R  K

1441 GCCTTTAAGAAATGGACACCTCCTCGGTCACCTTTTAATCTCGTTCAAGAAACACTTTTTCATGATCCATGGAAGCTTCTCATCGCTACT 1530
      A  F  K  K  W  T  P  P  R  S  P  F  N  L  V  Q  E  T  L  F  H  D  P  W  K  L  L  I  A  T

1531 ATATTTCTCAATCGGACCTCAGGCAAAATGGCAATACCTGTGCTTTGGAAGTTTCTGGAGAAGTATCCTTCAGCTGAGGTAGCAAGAACC 1620
      I  F  L  N  R  T  S  G  K  M  A  I  P  V  L  W  K  F  L  E  K  Y  P  S  A  E  V  A  R  T

1621 GCAGACTGGAGAGATGTGTCAGAACTTCTTAAACCTCTTGGTCTCTACGATCTTCGGGCAAAAACCATTGTCAAGTTCTCAGATGAATAC 1710
      A  D  W  R  D  V  S  E  L  L  K  P  L  G  L  Y  D  L  R  A  K  T  I  V  K  F  S  D  E  Y

1711 CTGACAAAGCAGTGGAAGTATCCAATTGAGCTTCATGGGATTGGTAAATATGGCAACGACTCTTACCGAATTTTTTGTGTCAATGAGTGG 1800
      L  T  K  Q  W  K  Y  P  I  E  L  H  G  I  G  K  Y  G  N  D  S  Y  R  I  F  C  V  N  E  W

1801 AAGCAGGTGCACCCTGAAGACCACAAATTAAATAAATATCATGACTGGCTTTGGGAAAATCATGAAAAATTAAGTTTATCTTAAACTCTG 1890
      K  Q  V  H  P  E  D  H  K  L  N  K  Y  H  D  W  L  W  E  N  H  E  K  L  S  L  S  *

1891 CAGCTTTCAAGCTCATCTGTTATGCATAGCTTTGCACTTCAAAAAAGCTTAATTAAGTACAACCAACCACCTTTCCAGCCATAGAGATTT 1980
1981 TAATTAGCCCAACTAGAAGCCTAGTGTGTGTGCTTTCTTAATGTGTGTGCCAATGGTGGATCTTTGCTACTGAATGTGTTTGAACATGTT 2070
2071 TTGAGATTTTTTTAAAATAAATTATTATTTGACAACAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2152
```

Fig. 3

```
hmed1   KEDVAMELER V......... GEDEEQMMIK RSSECNPLLQ EPIAS
rmecp2  KEDKEGKHEP LQPSAHHSAE PAEAGKAETS ESSGSAPAVP EASAS

hmed1   AQ.....FGA TAGTECRKSV PCGWERVVKQ RLFGKTAGRF DVYFI
rmecp2  PKQRRSIIRD RGPMYDDPTL PEGWTRKLKQ RKSGRSAGKY DVYLI

hmed1   SPQGLKFRSK SSLANYLHKN GETSLKPEDF DFTVLSKRGI KSR K
rmecp2  NPQGKAFRSK VELIAYFEKV GDTSLDPNDF DFTVTG.RGS PSR .

hmed1   EDKEGKHEP
rmecp2  .........
```

Fig. 4A

```
     a-hmed1  DPWKLLIATI  FLNRTSGKMA  IPVLWKFLEK  YPSAEVARTA  DWREVSELLK  PLGLYDLRAK
b-endo3ecoli  SPFELLIAVL  LSAQATDVSV  NKATAKFYPV  ANTPAAMLEL  GVEGVKTYIK  TIGLYNSKAE
    c-gtmrmt  DPYVILITEI  LLRRTTAGHV  KKIYDKFFVK  YKCFEDILKT  PKSEVAKDIK  EIGLSNQRAE
   d-uvendoml  TPFELLVATV  LSAQTTDVRV  NAATPAFFAR  FPDAHAMAAA  TEPELQELVR  STGLYRNKAS
 e-mutyecoli  TPYKVWLSEV  MLQQTQVATV  IPYFERFMAR  FPTVTDLANA  PLDEVLHLWT  GLGY.YARAR

     a-hmed1  TIVKFSDEYL  TKQ.......  .WKYPIELHG  VGKYGNDSY.  ..........  ...RIFCVNE
b-endo3ecoli  NIIKTCRILL  EQHNGEVPED  RAALEA.LPG  VGRKTANVVL  NTAFGWPTIA  VDTHIFRVCN
    c-gtmrmt  QLKELARVVI  NDYGGRVPRN  RKAILD.LPG  VGKYTCAAVM  CLAFGKKAAM  VDANFVRVIN
   d-uvendoml  AILRLSQELV  GRHDGEVPAR  LEDLVA.LPG  VGRKTAFVVL  GNAFGQPGIT  VDTHFGRLAR
 e-mutyecoli  NLHKAAQQVA  TLHGGKFPET  FEEVAA.LPG  VGRSTAGAVL  SLSAGKHFPI  LDGNVKRVLA

     a-hmed1  WKQVHPEDHK  LNKYHDWLWE  NHEKLSLS..  ..........  ..........  ..........
b-endo3ecoli  RTQFAPGKNV  EQVEEKLLKV  VPA..EFKVD  CHHW...LIL  HGRYTCIARK  PRCGSCIIED
    c-gtmrmt  RYFG.GSYEN  LNYNHKALWE  LAETLVPGGK  CRDFNLGLMD  FSAIICAPRK  PKCEKCGMSK
   d-uvendoml  RLGFTDETDP  GKGRARRGWP  VPPARDWTML  S.HR...LIF  HGRRVCHARR  PACGRCPIAR
 e-mutyecoli  RCYAVSGWPG  KKEVENKLWS  LSEQVTPAVG  VERFNQALMD  LGAMICTRSK  PKCSLCPLQN

     a-hmed1  ..........
b-endo3ecoli  LCEYKEKVDI
    c-gtmrmt  LCSYYEKCST
   d-uvendoml  WCPSYAAGET
 e-mutyecoli  GCIAAANNSW
```

Fig. 4B

```
MED1 IASAQFGATA GTECRKSVPC GWERVVKQRL FGKTAGRFDV YFISPQGLKF RSKSSLANYL 60
PCM1 MAEDWLDCPA .......LGP GWKRREVFRK SGATCGRSDT YYQSPTGDRI RSKVELTRYL 53

MED1 HKNGETSLKP EDFDFTVLSK RGIKS85
PCM1 GPACDLT..L FDFKQGILCY PAPKA76
```

Fig. 4C

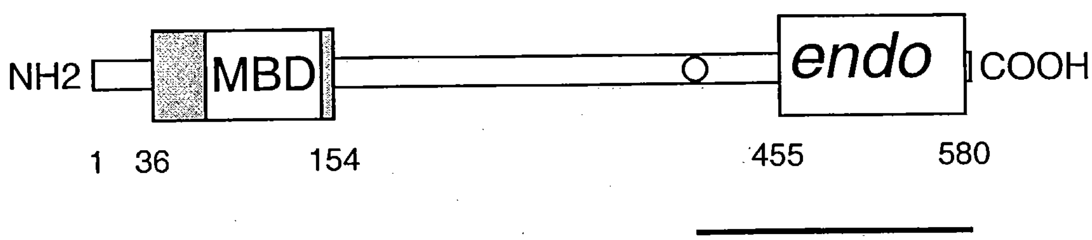


Fig. 5

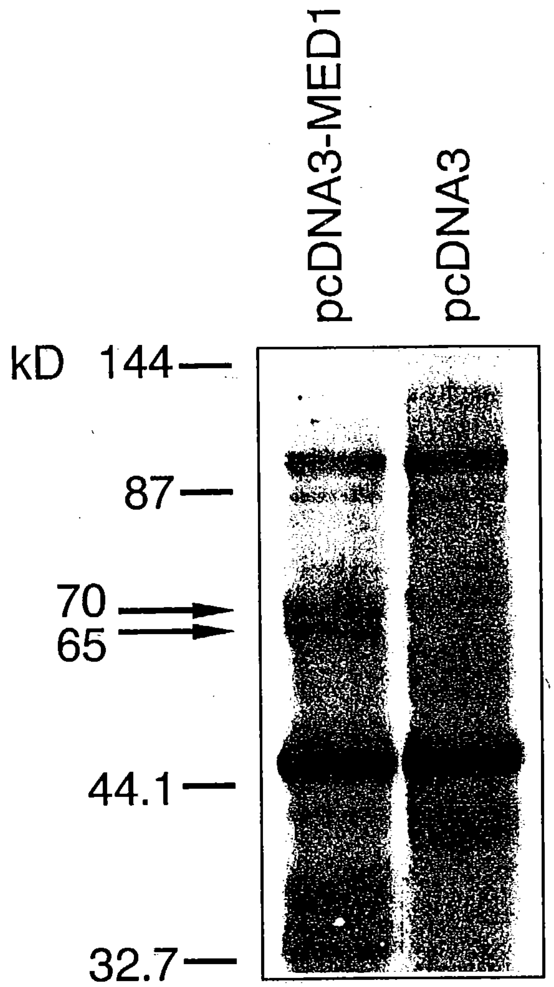
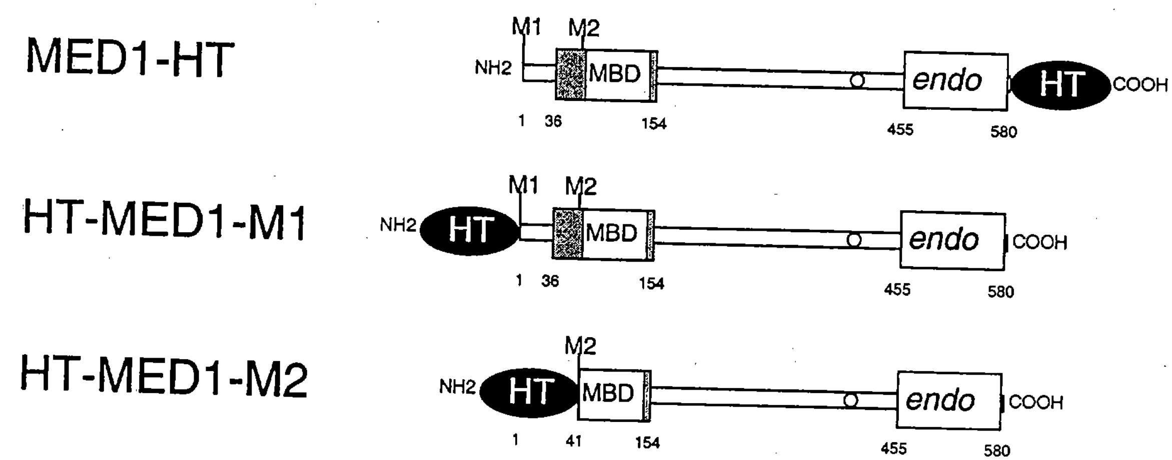


Fig. 6

Fig. 7A



MED1-HT
HT-MED1-M1
HT-MED1-M2
M1
M2
NH2
MBD
endo
HT
COOH
1
36
41
154
455
580

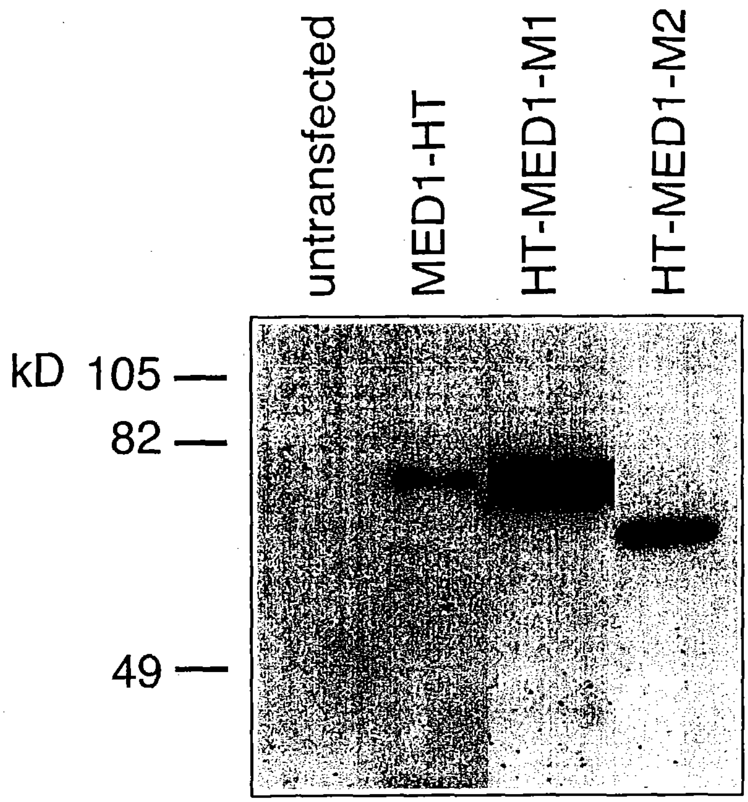


Fig. 7B

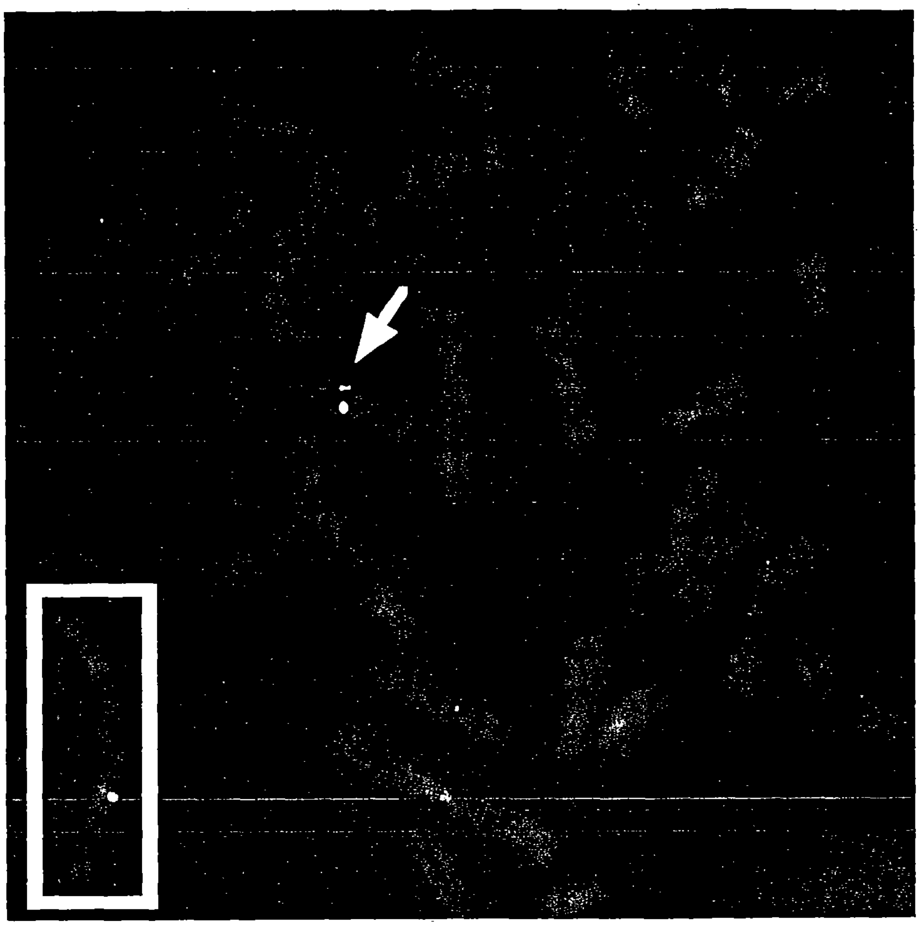


Fig. 8

Fig. 9A

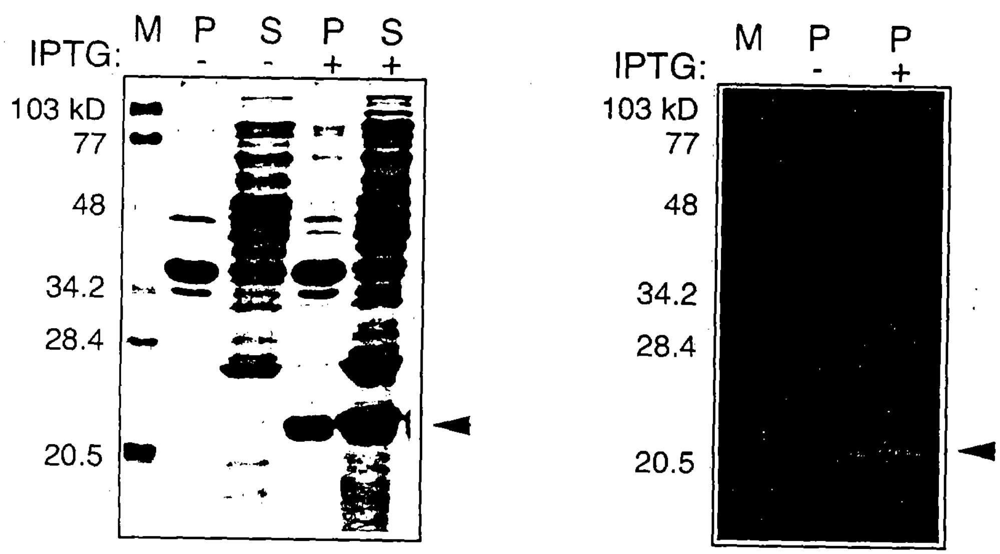
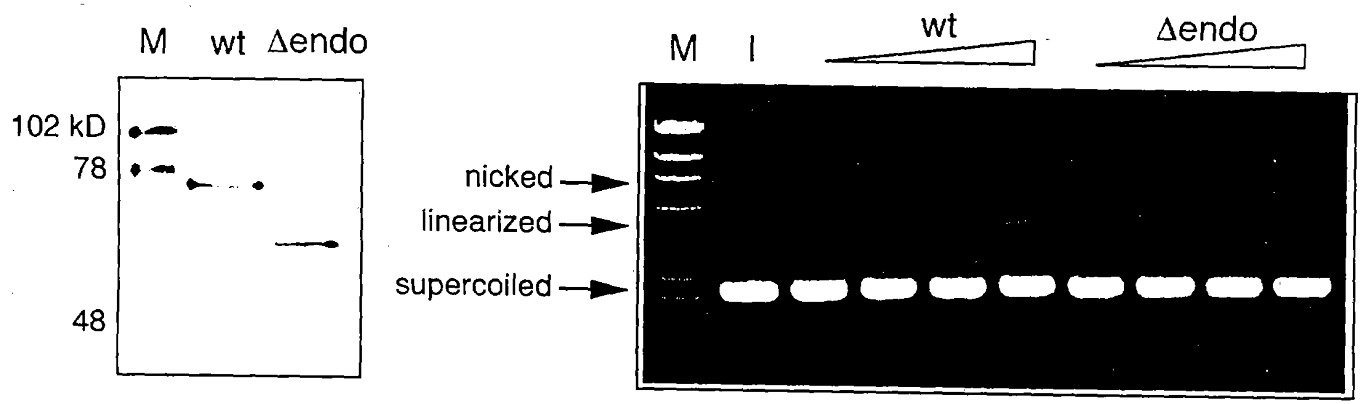

IPTG:
M
P
S
P
S
-
-
+
+
103 kD
77
48
34.2
28.4
20.5
IPTG:
M
P
P
-
+
103 kD
77
48
34.2
28.4
20.5

Fig. 9B


M
wt
Δendo
102 kD
78
48
M
I
wt
Δendo
nicked
linearized
supercoiled

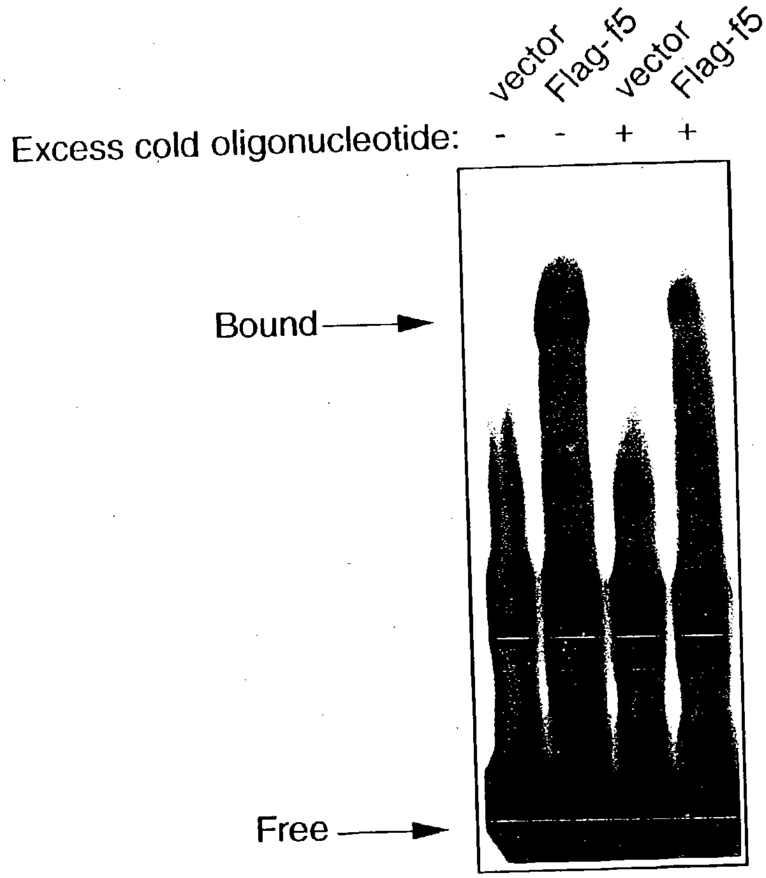

vector
Flag-f5
vector
Flag-f5
Excess cold oligonucleotide: - - + +
Bound
Free

Fig. 10A

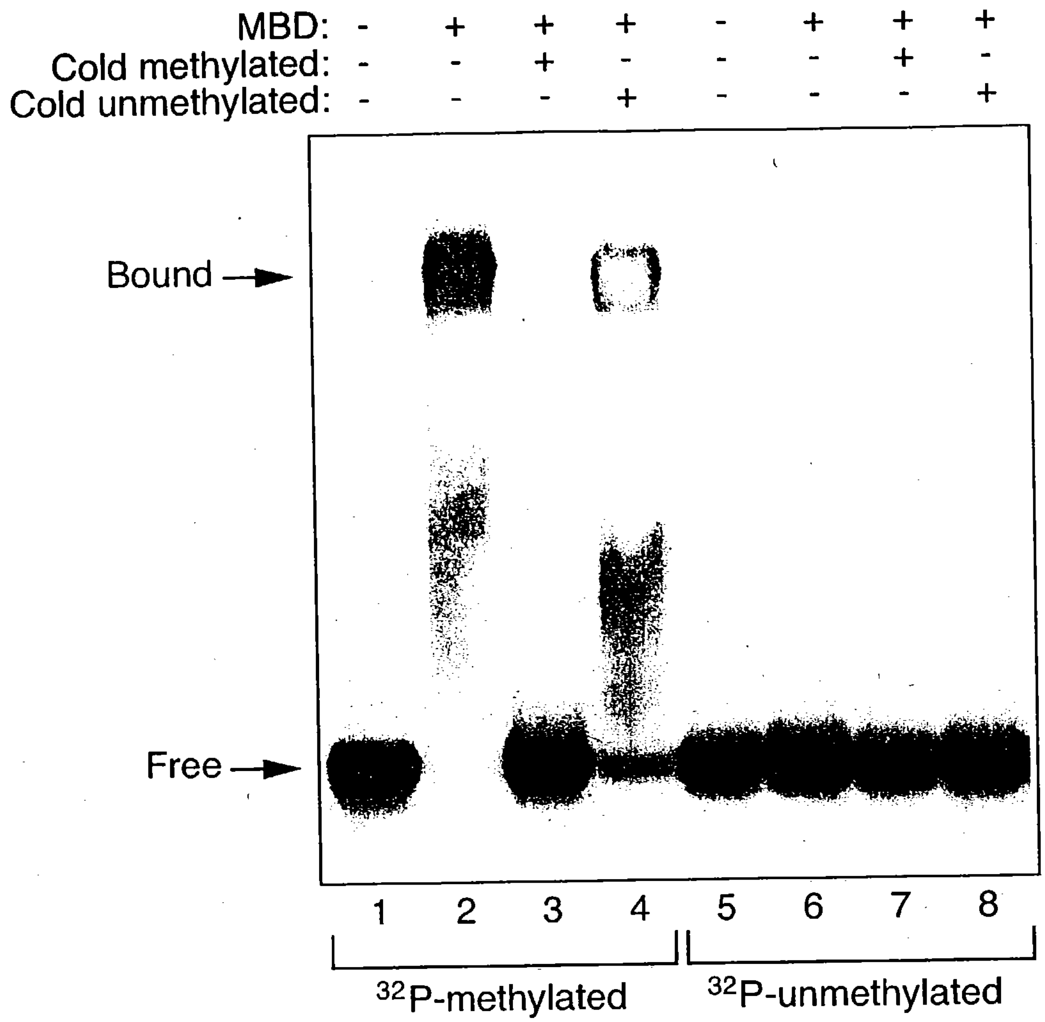

MBD: - + + + - + + +
Cold methylated: - - + - - - + -
Cold unmethylated: - - - + - - - +
Bound
Free
1 2 3 4 5 6 7 8
32P-methylated
32P-unmethylated

Fig. 10B

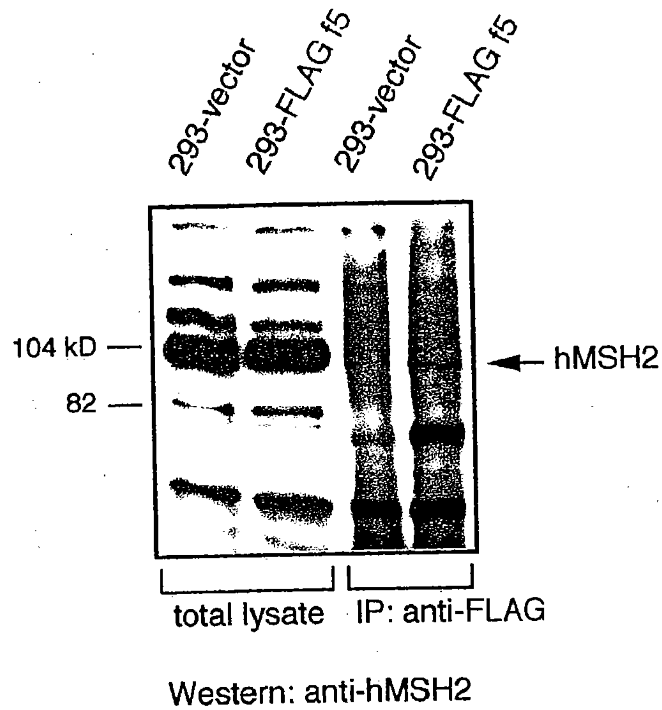

293-vector
293-FLAG f5
293-vector
293-FLAG f5
104 kD
82
hMSH2
total lysate
IP: anti-FLAG
Western: anti-hMSH2

Fig. 11A

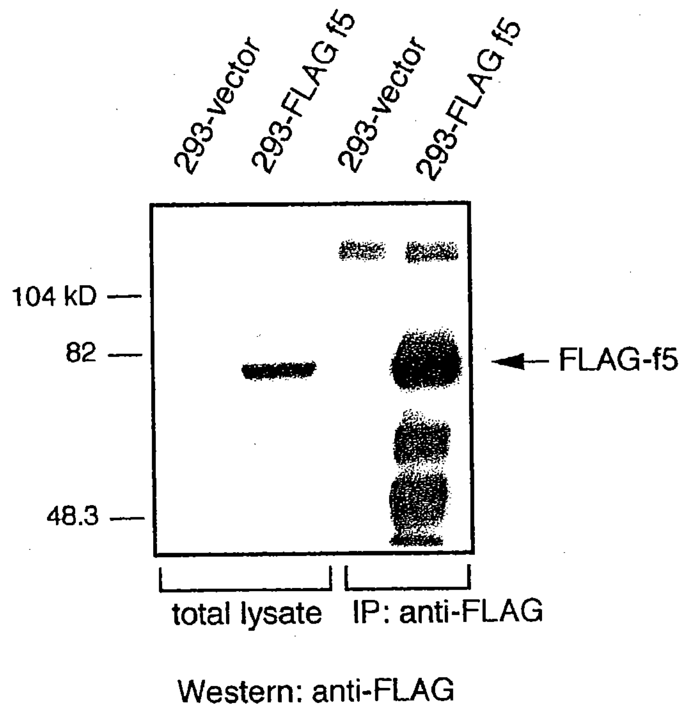

293-vector
293-FLAG f5
293-vector
293-FLAG f5
104 kD
82
48.3
FLAG-f5
total lysate
IP: anti-FLAG
Western: anti-FLAG

Fig. 11B

Fig. 11C

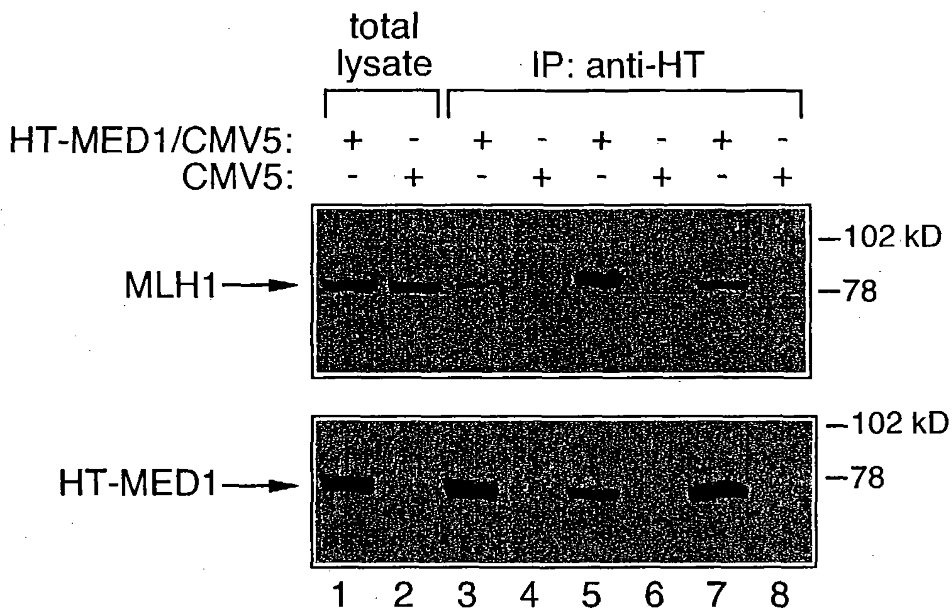

total
lysate
IP: anti-HT
HT-MED1/CMV5: + - + - + - + -
CMV5: - + - + - + - +
MLH1
—102 kD
—78
HT-MED1
—102 kD
—78
1 2 3 4 5 6 7 8

Fig. 12A

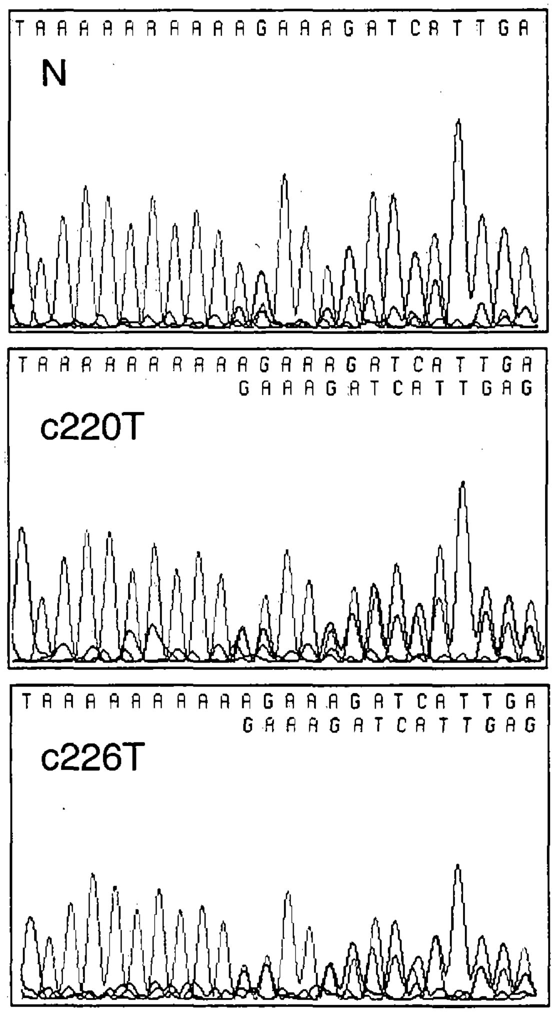


Fig. 12B

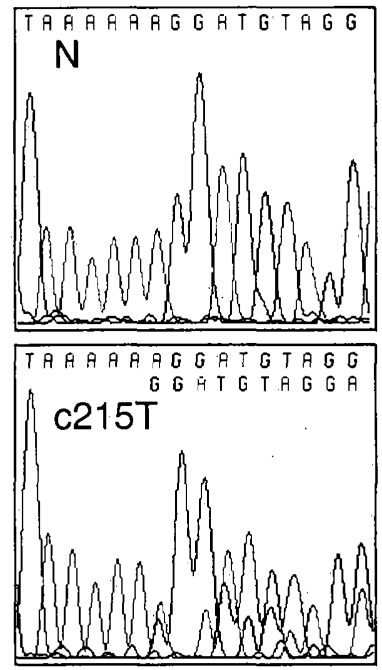
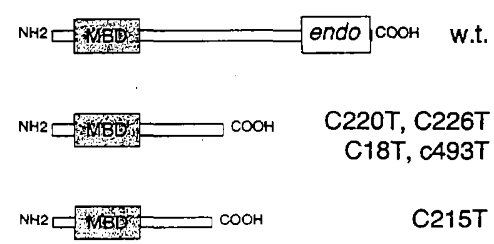


Fig. 12C



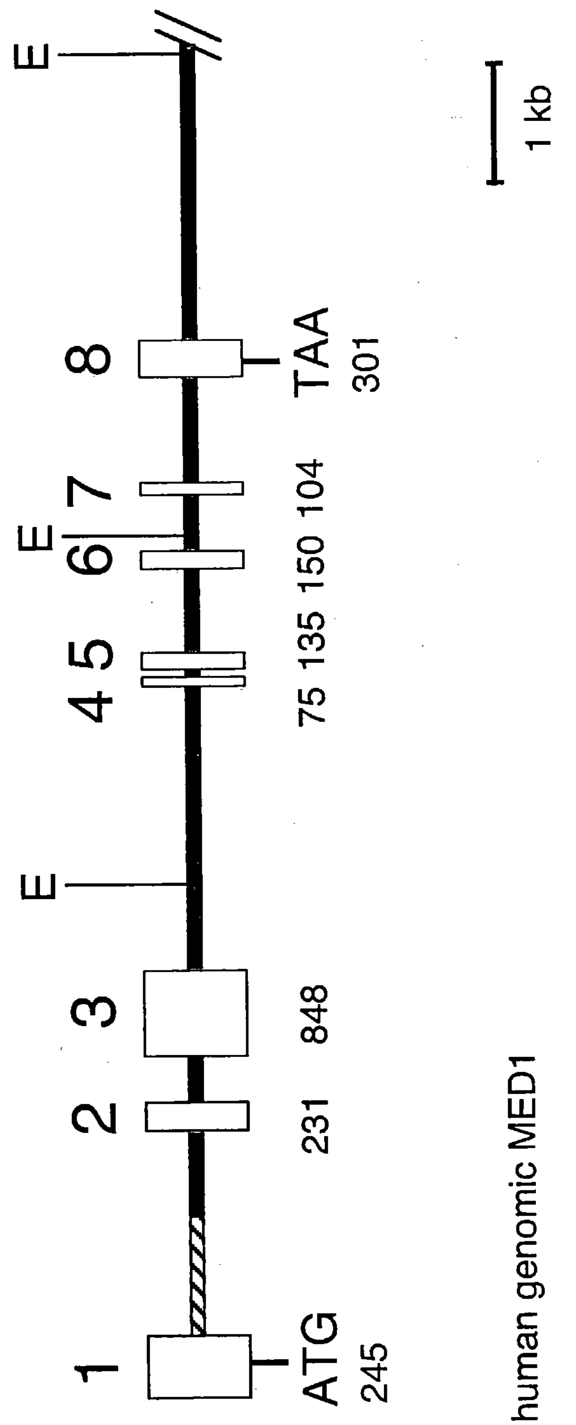

E
E
E
1
ATG
245
2
231
3
848
4
5
75
135
6
150
7
104
8
TAA
301
1 kb
human genomic MED1

Fig. 13

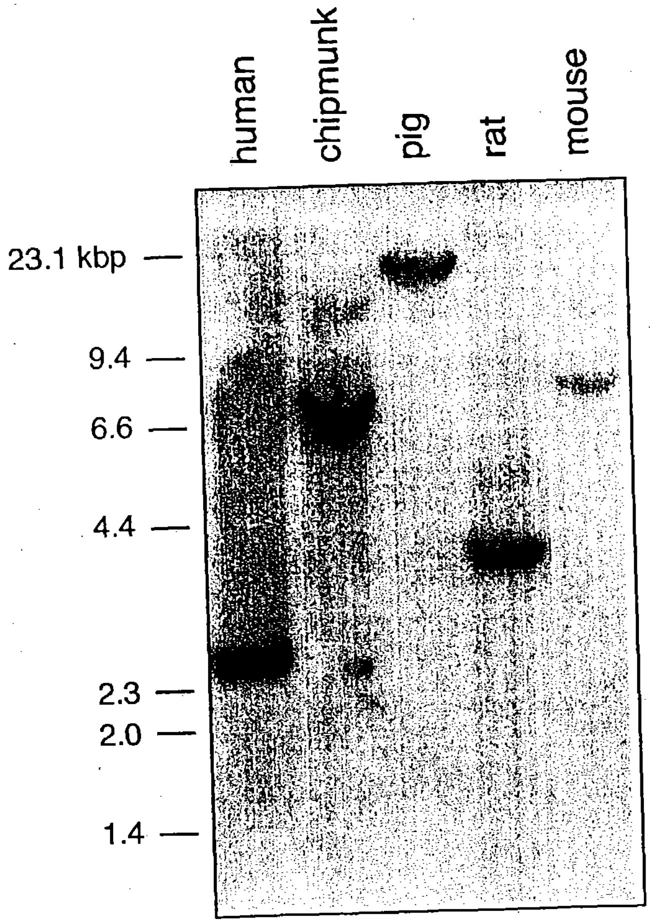


Fig. 14A

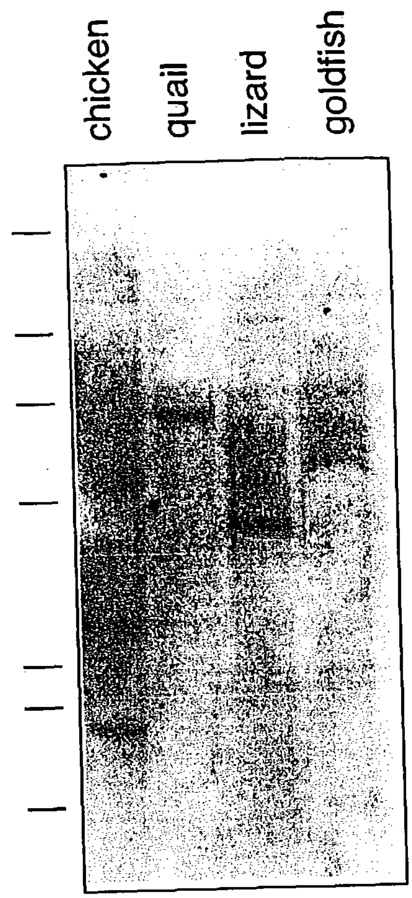


Fig. 14B

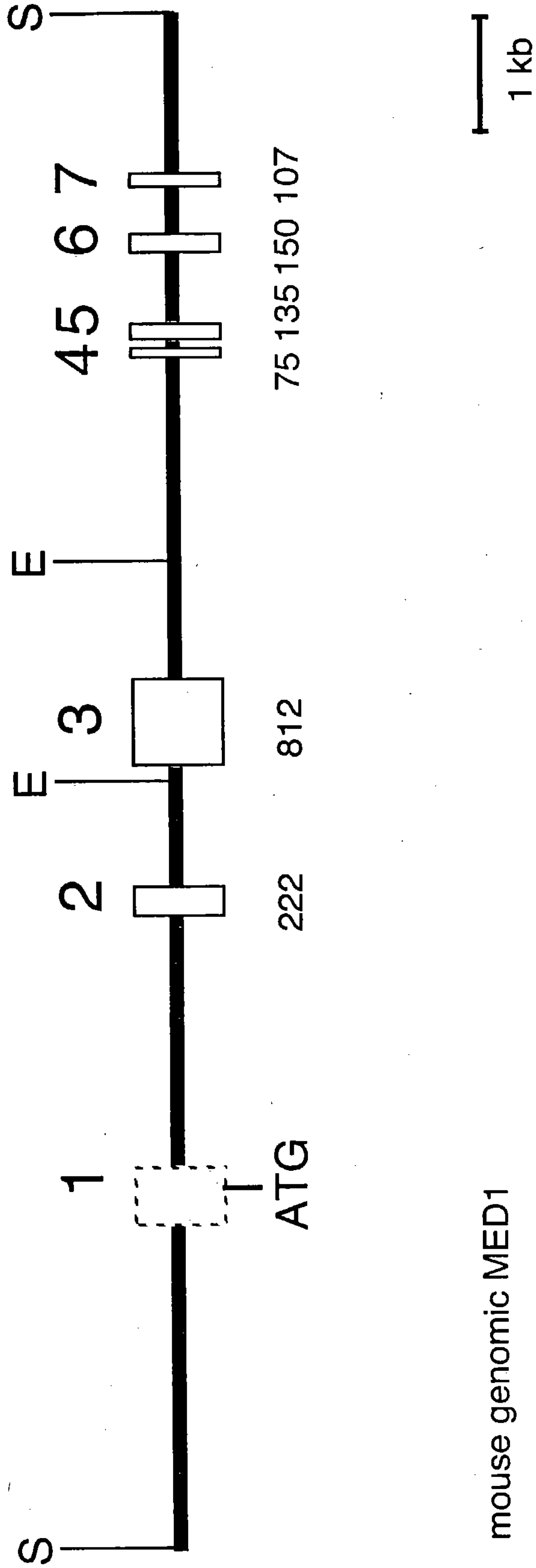

S
1
2
3
45
6
7
E
E
S
ATG
222
812
75 135 150 107
1 kb
mouse genomic MED1

Fig. 15

1 CAAGGAAGAT ATTGCTGTTG GACTGGGAGG AGTGGGAGAA GATGGAAAGG

51 ACCTGGTGAT AAGCAGTGAG CGCAGCTCCC TTCTCCAAGA GCCCACTGCT

101 TCTACTCTGT CTAGTACTAC AGCGACAGAA GGCCACAAGC CTGTCCCGTG

151 TGGATGGGAA AGAGTTGTGA AGCAAAGGTT ATCTGGGAAA ACTGCAGGAA

201 AATTTGATGT ATACTTTATC AGCCCACAAG GATTGAAGTT CAGATCAAAA

251 CGTTCACTTG CTAATTATCT TCTCAAAAAT GGGGAGACTT TTCTTAAGCC

301 TGAAGATTTT AATTTTACTG TACTGCCGAA AGGGAGCATC AATCCCGGTT

351 ATAAACACCA AAGTTTGGCA GCTCTGACTT CCCTGCAGCC AAATGAAACT

401 GACGTTTCAA AGCAGAACCT CAAGACACGA AGCAAGTGGA AAACAGATGT

451 GTTGCCTCTG CCCAGTGGTA CTTCAGAGTC GCCAGAAAGC AGCGGACTGT

501 CTAACTCTAA CTCGGCTTGC TTGCTATTGA GAGAACATAG GGACATTCAG

551 GATGTTGACT CTGAGAAGAG GAGAAAGTCC AAAAGAAAGG TGACTGTTTT

601 GAAAGGAACT GCAAGTCAGA AAACCAAACA AAAGTGCAGG AAGAGTCTCT

651 TAGAGTCTAC TCAAAGAAAC AGAAAAAGAG CATCTGTGGT TCAGAAGGTG

701 GGTGCTGATC GCGAGCTGGT GCCACAGGAA AGTCAACTCA ACAGAACCCT

751 CTGCCCTGCA GATGCCTGTG CAAGGGAGAC TGTTGGCCTG GCTGGGGAAG

801 AAAAATCACC AAGCCCAGGA CTGGATCTTT GTTTCATACA AGTAACTTCT

851 GGCACCACAA ACAAATTCCA TTCAACTGAA GCAGCAGGTG AAGCAAATCG

901 TGAGCAGACT TTTTTAGAAT CAGAGGAAAT CAGATCGAAG GGAGACAGAA

951 AGGGGGAGGC ACATTTGCAT ACTGGTGTTT TACAGGATGG CTCTGAAATG

1001 CCCAGCTGCT CACAAGCCAA GAAACACTTT ACTTCTGAGA CATTTCAAGA

1051 AGACAGCATC CCACGGACAC AAGTAGAAAA AAGGAAAACA AGCCTGTATT

1101 TTTCCAGCAA GTACAACAAA GAAGCTCTTA GCCCCCCAAG ACGCAAATCC

1151 TTCAAGAAAT GGACCCCTCC TCGGTCACCT TTTAATCTTG TTCAAGAAAT

1201 ACTTTTCCAT GACCCATGGA AGCTCCTCAT CGCGACTATA TTTCTCAATC

1251 GGACCTCAGG CAAGATGGCC ATCCCTGTGC TGTGGGAGTT TCTAGAGAAG

1301 TACCCTTCAG CTGAAGTGGC CCGAGCTGCC GACTGGAGGG ACGTGTCGGA

Fig. 16A

1351 GCTTCTCAAG CCTCTTGGTC TCTACGATCT CCGTGCAAAA ACCATTATCA

1401 AGTTCTCAGA TGAATATCTG ACAAAGCAGT GGAGGTATCC GATTGAGCTT

1451 CATGGGATTT GGTTAAAATA TGGCAACGAC TCTACCGGAT CTTTTGTGTC

1501 AATGAATGGA ACAG

Fig. 16B

mouse MED1 protein (upper sequence) x human MED1 protein (lower sequence)

```
  1 KEDIAVGLGGVGEDGKDLVI..SSERSSLLQEPTAST.LSSTTATEGHKP 47

 36 KEDVAMELERVGEDEEQMMIKRSSECNPLLQEPIASAQFGATAGTECRKS 85

 48 VPCGWERVVKQRLSGKTAGKFDVYFISPQGLKFRSKRSLANYLLKNGETF 97

 86 VPCGWERVVKQRLFGKTAGRFDVYFISPQGLKFRSKSSLANYLHKNGETS 135

 98 LKPEDFNFTVLPKGSINPGYKHQSLAALTSLQPNETDVSKQNLKTRSKWK 147

136 LKPEDFDFTVLSKRGIKSRYKDCSMAALTSHLQNQSNNSNWNLRTRSKCK 185

148 TDVLPLPSGTSESPESSGLSNSNSACLLLREHRDIQDVDSEKRRKSKRKV 197

186 KDVFMPPSSSSELQESRGLSNFTSTHLLLKEDEGVDDVNFRKVRKPKGKV 235

198 TVLKGTASQKTKQKCRKSLLESTQRNRKRAS................... 228

236 TILKGIPIKKTKKGCRKSCSGFVQSDSKRESVCNKADAESEPVAQKSQLD 285

                          .
                          .
                          .

229 ...............EDSIPRTQVEKRKTSLYFSSKYNKEALSPPRRKSF 263

386 CSPTRKDFTGEKIFQEDTIPRTQIERRKTSLYFSSKYNKEALSPPRRKAF 435

264 KKWTPPRSPFNLVQEILFHDPWKLLIATIFLNRTSGKMAIPVLWEFLELY 313

436 KKWTPPRSPFNLVQETLFHDPWKLLIATIFLNRTSGKMAIPVLWKFLEKY 485

314 PSAEVARAADWRDVSELLKPLGLYDLRAKTIIKFSDEYLTKQWRYPIELH 363

486 PSAEVARTADWRDVSELLKPLGLYDLRAKTIVKFSDEYLTKQWKYPIELH 535

364 GIWLKYGNDSYRIFCVNEWKQ 384

536 GIG.KYGNDSYRIFCVNEWKQ 555
```

Fig. 17

Fig. 18A

Exon 2

ggttttgtttttcc**ag**CAAGGAAGATATTGCTGTTGGACTGGGAGGAGTG
GGAGAAGATGGAAAGGACCTGGTGATAAGCAGTGAGCGCAGCTCCCTTCT
CCAAGAGCCCACTGCTTCTACTCTGTCTAGTACTACAGCGACAGAAGGCC
ACAAGCCTGTCCCGTGTGGATGGGAAAGAGTTGTGAAGCAAAGGTTATCT
GGGAAAACTGCAGGAAAATTTGATGTATACTTTATCAG**gt**aagcatttag
Gaaggaaaata

Fig. 18B

Exon 3

ctttttttttttcctttta**ag**CCCACAAGGATTGAAGTTCAGATCAAAAC
GTTCACTTGCTAATTATCTTCTCAAAAATGGGGAGACTTTTCTTAAGCCT
GAAGATTTTAATTTTACTGTACTGCCGAAAGGGAGCATCAATCCCGGTTA
TAAACACCAAAGTTTGGCAGCTCTGACTTCCCTGCAGCCAAATGAAACTG
ACGTTTCAAAGCAGAACCTCAAGACACGAAGCAAGTGGAAAACAGATGTG
TTGCCTCTGCCCAGTGGTACTTCAGAGTCGCCAGAAAGCAGCGGACTGTC
TAACTCTAACTCGGCTTGCTTGCTATTGAGAGAACATAGGGACATTCAGG
ATGTTGACTCTGAGAAGAGGAGAAAGTCCAAAAGAAAGGTGACTGTTTTG
AAAGGAACTGCAAGTCAGAAAACCAAACAAAAGTGCAGGAAGAGTCTCTT
AGAGTCTACTCAAAGAAACAGAAAAAGAGCATCTGTGGTTCAGAAGGTGG
GTGCTGATCGCGAGCTGGTGCCACAGGAAAGTCAACTCAACAGAACCCTC
TGCCCTGCAGATGCCTGTGCAAGGGAGACTGTTGGCCTGGCTGGGGAAGA
AAAATCACCAAGCCCAGGACTGGATCTTTGTTTCATACAAGTAACTTCTG
GCACCACAAACAAATTCCATTCAACTGAAGCAGCAGGTGAAGCAAATCGT
GAGCAGACTTTTTTAGAATCAGAGGAAATCAGATCGAAGGGAGACAGAAA
GGGGGAGGCACATTTGCATACTGGTGTTTTACAGGATGGCTCTGAAATGC
CCAGCTGCTCACAAGCCAAGAAACACTTTACTTCTGAGACATTTCAAG**gt**
actcagtgcatgaaaa

Fig. 18C

Exon 4

gactataaactaattttgcttctc**ag**AAGACAGCATCCCACGGACACAAG
TAGAAAAAAGGAAAACAAGCCTGTATTTTTCCAGCAAGTACAACAAAGAA
G**gt**acccacctttccctaagc

Fig. 18D

Exon 5

tatatttntgn**ag**CTCTTAGCCCCCCAAGACGCAAATCCTTCAAGAAATG
GACCCCTCCTCGGTCACCTTTTAATCTTGTTCAAGAAATACTTTTCCATG
ACCCATGGAAGCTCCTCATCGCGACTATATTTCTCAATCGGACCTCAG**g**
ttnggggtcattgncat

Fig. 18E

Exon 6

tgtttatgctcccc**ag**GCAAGATGGCCATCCCTGTGCTGTGGGAGTTTCT
AGAGAAGTACCCTTCAGCTGAAGTGGCCCGAGCTGCCGACTGGAGGGACG
TGTCGGAGCTTCTCAAGCCTCTTGGTCTCTACGATCTCCGTGCAAAAACC
ATTATCAAGTTCTCAG**gt**atgtccccagcccag

Fig. 18F

Exon 7

tggatgtgtatccctc**ag**ATGAATATCTGACAAAGCAGTGGAGGTATCCG
ATTGAGCTTCATGGGATTTGGTTAAAATATGGCAACGACTCTACCGGAT
CTTTTGTGTCAATGAATGGAACAG**gt**aagcccaccactggggcc

Fig. 19A

Exon 1

GCGGCGGCGTCTGGGGCGCTTTCGCAACATTCAGACCTCGGTTGCAGCCCGGTGCCGTGAGCTGAA
GAGGTTTCACATCTTACTCCGCCCCACACCCTGGGCGTTGCGGCGCTGGGCTCGTTGCTGCAGCCG
GACCCTGCTCGATGGGCACGACTGGGCTGGAGAGTCTGAGTCTGGGGGACCGCGGAGCTGCCCCCA
CCGTCACCTCTAGTGAGCGCCTAGTCCCAGACCCGCCGAATGACCTCCG**gt**aagttactgtcccct
tttgggcctcagtttcaccacctgtaaaatggtatcgggagagtggacagtgtgtgggcctttcta
acctttgacagagggtcggcanaaacctcgaagcccacgggtttagttactagggtctggagccca
ggtgctcttcctgtgcgatcagc...

Fig. 19B

Exon 2

...tttggaagacaggaaat<u>**actcccatagcacaagactgg**</u>tccacactgactttaatctccc
tcattttaatatggataatctatgtggttcctgcattgtcatggattaaaactgagtaggcagtgg
aagataaattttaaataagttaatcacttagactttgtttttcc**ag**CAAAGAAGATGTTGCTATGG
AATTGGAAAGAGTGGGAGAAGATGAGGAACAAATGATGATAAAAAGAAGCAGTGAATGTAATCCCT
TGCTACAAGAACCCATCGCTTCTGCTCAGTTTGGTGCTACTGCAGGAACAGAATGCCGTAAGTCTG
TCCCATGTGGATGGGAAAGAGTTGTGAAGCAAAGGTTATTTGGGAAGACAGCAGGAAGATTTGATG
TGTACTTTATCAG**gt**aagcatataagatggtaaagatagtacagccaaatgattttgtctgg<u>**gcag**</u>
<u>**gtagtgggagcatagc**</u>aggaatcttagcttctttatattttaccataaaccattgcagattc
tattctttcaatgttgctattaattacatcaagtgatttggggaaattacatacattttgtccct
ccttctgtgaatggttaacgggtaggttgcattttagttatatttataaatttatattgtcataga
ggaaccatttaaaggccattatcactctttttcatttttaaatgacagagacctatggcaacatt
tggaaattaattagaatctgaaatgtggtccagttcttttaaaagtcccttctatttactagcagt
aagtttcctttaatatcattttct**ag(continues into exon 3, see below)**

Fig. 19C

Exon 3

aatctgaaatgtggtccagttcttttaaaagt**cccttctatttactagcagta**agtttccttt
aatatcattttct**ag**CCCACAAGGACTGAAGTTCAGATCCAAAAGTTCACTTGCTAATTATCTTCA
CAAAAATGGAGAGACTTCTCTTAAGCCAGAAGATTTTGATTTTACTGTACTTTCTAAAAGGGGTAT
CAAGTCAAGATATAAAGACTGCAGCATGGCAGCCCTGACATCCCATCTACAAAACCAAAGTAACAA
TTCAAACTGGAACCTCAGGACCCGAAGCAAGTGCAAAAAGGATGTGTTTATGCCGCCAAGTAGTAG
TTCAGAGTTGCAGGAGAGCAGAGGACTCTCTAACTTTACTTCCACTCATTTGCTTTTGAAAGAAGA
TGAGGGTGTTGATGATGTTAACTTCAGAAAGGTTAGAAAGCCCAAAGGAAAGGTGACTATTTTGAA
AGGAATCCCAATTAAGAAAACTAAAAAAGGATGTAGGAAGAGCTGTTCAGGTTTTGTTCAAAGTGA
TAGCAAAAGANAATCTGTGTGTAATAAAGCAGATGCTGAAAGTGAACCTGTTGCACAAAAAAGTCA
GCTTGATAGAACTGTCTGCATTTCTGATGCTGGAGCATGTGGTGAGACCCTCAGTGTGAGCAGTGA
AGAAAACNGCCTTGTAAAAAAAAAAGAAAGATCATTGAGTTCAGGATCAAATTTTTGTTCTGAACA
AAAAACTTCTGGCATCATAAACAAATTTTGTTCAGCCAAAGACTCAGAACACAACGAGAAGTATGA
GGATACCTTTTTAGAATCTGAAGAAATCGGAACAAAAGTAGAAGTTGTGGAAAGGAAAGAACATTT
GCATACTGACATTTTAAAACGTGGCTCTGAAATGGACAACAACTGCTCACCAACCAGGAAAGACTT
CACTG**gt**gagaaaatatttcaag**gt**atccagtgctttcagcactattaaacattagtgat**gagaa**
atttatatgctgcatctgtatcgtgccatac

Please note: at the end of exon 3, two alternative splice donor sites are present (see Sequence Variations, page 40 of the application).

Fig. 19D

Exon 4 and Exon 5

tagtaccaagttcatgggtcattagttagattaattgggtatttatgtaaagggcttagaatagtg
cctggcatgctttgtaatagtgttgatattattatt**tgcatccctcaatattgcttt**aagcta
aaccatagactccataaagtgtttacttttccttttc**ag**AAGATACCATCCCACGAACACAGATAG
AAAGAAGGAAAACAAGCCTGTATTTTTCCAGCAAATATAACAAAGAAG**gt**atccctttcccaatca
gaacagcaaattctaattccattttgggttttcaattctgatgcactatgtttgttt**ag**CTCTTAG
CCCCCCACGACGTAAAGCCTTTAAGAAATGGACACCTCCTCGGTCACCTTTTAATCTCGTTCAAGA
AACACTTTTTCATGATCCATGGAAGCTTCTCATCGCTACTATATTTCTCAATCGGACCTCAG**gt**tt
ggggattattatcatctttgtcttagtagagacagtgtggt**agggagaaagcactgaattga**g
gcctgggttcaaagtcattttgagtgtgtcacctgggatagggcattccccctttcacccttaaac
tcttcacctatgaggaaaatggggg

Fig. 19E

Exon 6

ccagtgtttttgtttttgttttctttaaaaaaaaaaaaaaccctctggatgagatttctatga
gaaactacttgaacgtgaaatcagcccacctggagtcttgtaatcattcagttacttttacnttcc
c**ag**GCAAAATGGCAATACCTGTGCTTTGGAAGTTTCTGGAGAAGTATCCTTCAGCTGAGGTAGCAA
GAACCGCAGACTGGAGAGATGTGTCAGAACTTCTTAAACCTCTTGGTCTCTACGATCTTCGGGCAA
AAACCATTGTCAAGTTCTCAG**gt**attttcctatacacccaaaggaaaaacataatacattgtgctt
atttaa**gagagccacaccttaaacttt**aatgttctcagatactatattaatggaggtttttca
gctcaagcatttaaaaaagtccacttttccccaaaccacagtctcccactgacctaaacaataaat
cttt

Fig. 19F

Exon 7

cttta**gaagctgacctgataatgtgg**gatgttgtattcttc**ag**ATGAATACCTGACAAAGCAG
TGGAAGTATCCAATTGAGCTTCATGGGATTGGTAAATATGGCAACGACTCTTACCGAATTTTTTGT
GTCAATGAGTGGAAGCAG**gt**gaggctcactcccatccataattcagcacatt**tggtctctgagg**
caaaataagtccaccattatggttaagacnatttattggggatacaaatgctattacagtcacaa
caattgtgttcctggctgcggggaagcgngtggcatgtgggttttggggttttgatcagtaggcg
ctcccagg

Fig. 19G

Exon 8

tgtgtgagattaccttaatataaggtataacttaaaatattcatgaatcccaggaggttaaaggtt
ataacttttaggtatgg**tatcgtaatgtactgtccccc**agcaaacatttaaaaagccaatttt
aaaaaatgtatttctgactaagttacatntaaggtctctgcctctgtatcttatgtttcttcc**ag**G
TGCACCCTGAAGACCACAAATTAAATAAATATCATGACTGGCTTTCCCAAAATCATGAAAAATTAA
GTTTATCTTAAACTCTGCAGCTTTCAAGCTCATCTGTTATGCATAGCTTTGCACTTCAAAAAAGCT
TAATTAAGTACAACCAACCACCTTTCCAGCCATAGAGATTTTAATTAGCCCAACTAGAAGCCTAGT
GTGTGTGCTTTCTTAATGTGTGTGCCAATGGTGGATCTTTGCTACTGAATGTGTTTGAACATGTTT
TGAGATTTTTTTAAAATAAATTATTATTTGACAACA*atccaaaaaaaatacggcttttccaatga
tgaaatataatcagaagatgaaaaatagttctaaactatcaataatacaaagcaaat**ttctatca**
gccttgctaaagctaggggcccactaaatatttt

Please note: asterisk indicates the poly(A) addition site.

Fig. 19H

Complete sequence of the intron between exon 7 and exon 8.

GGAAGCAGgtgaggctcactcccatccataattcagcacatttggtctctgaggcaaaataagtcc
accattatggttaagactatttattggatacaaatgctattacagtcacaaacaattgtgttcctg
gctgcggggaagcgagtggcatgtgggttttggggttttgatcagtaagcgctcccaagtccaca
aagaccagtccagcggcgtggcctctgactcatctccagtggtttgtcacctctggccctgttcct
gtcattccctatttgtgtgctatctctaagcctgacgtggttttcctcctgtcaaaagtacaccac
tacaggaaagcaggaaggtttgggccttgcaatgtatgcatattgggtttctcttagtggtctcag
actacgtttgtggtgactgggtcctgcttcagccctgttgaatatgcccagcctgtggcatgctgg
tggtcatcctggcagctggtgggtggcctggtatgctgcccactcagcttgagactcaccctcatg
cattcagccagtaggtctggccaagcctgaactgaaggaccatggtcctatcccagcttcatcaca
gcaatccattgtgacctgagaatccatttaacctctcggtctagaacctccttctggaaagtgagg
tattaatacttgactcaatgttatcgccaccccacattctaagtcatggttgagtagtaatttgga
cagtaccttgtaaattgtgtgagattaccttaatataaggtataacttaaaatattcatgaatccc
aggaggttaaaggttataacttttaggtatggtatcgtaatgtactgtcccccagcaaacatttaa
aaagccaattttaaaaaatgtatttctgactaagttacattaaggtctctgcctctgtatcttatg
tttcttccagGTGCACCC

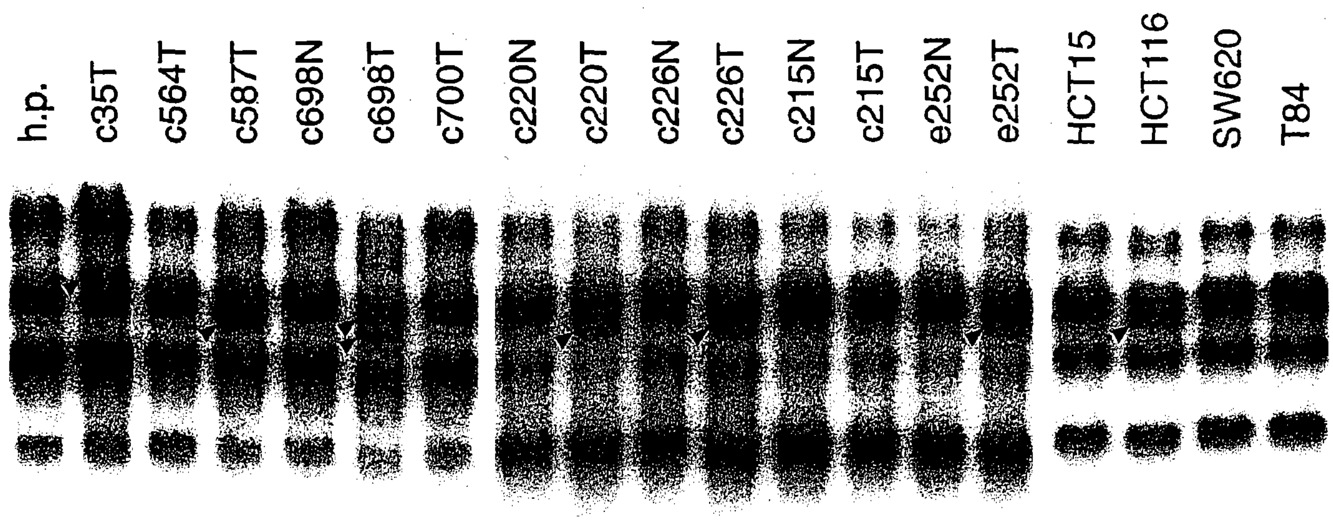

h.p.
c35T
c564T
c587T
c698N
c698T
c700T
c220N
c220T
c226N
c226T
c215N
c215T
e252N
e252T
HCT15
HCT116
SW620
T84

Fig. 20A

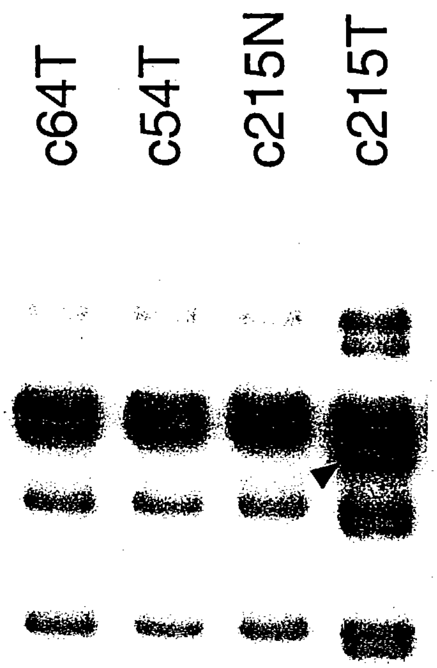


Fig. 20B

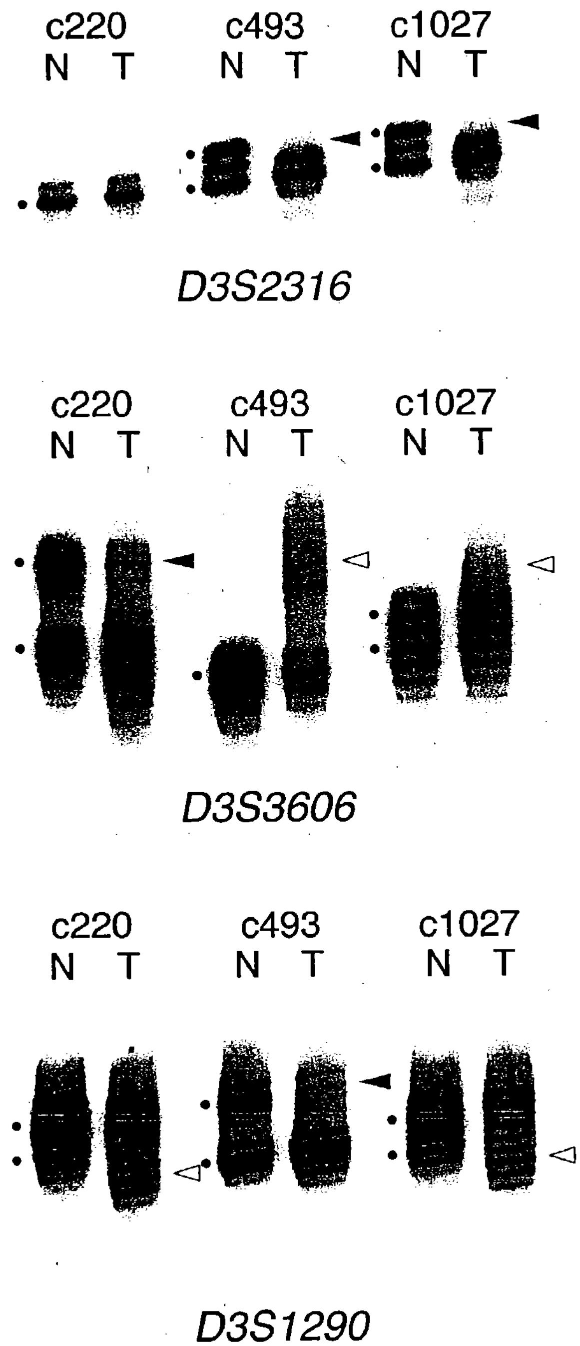

c220 c493 c1027
N T N T N T
D3S2316
c220 c493 c1027
N T N T N T
D3S3606
c220 c493 c1027
N T N T N T
D3S1290

Fig. 20C

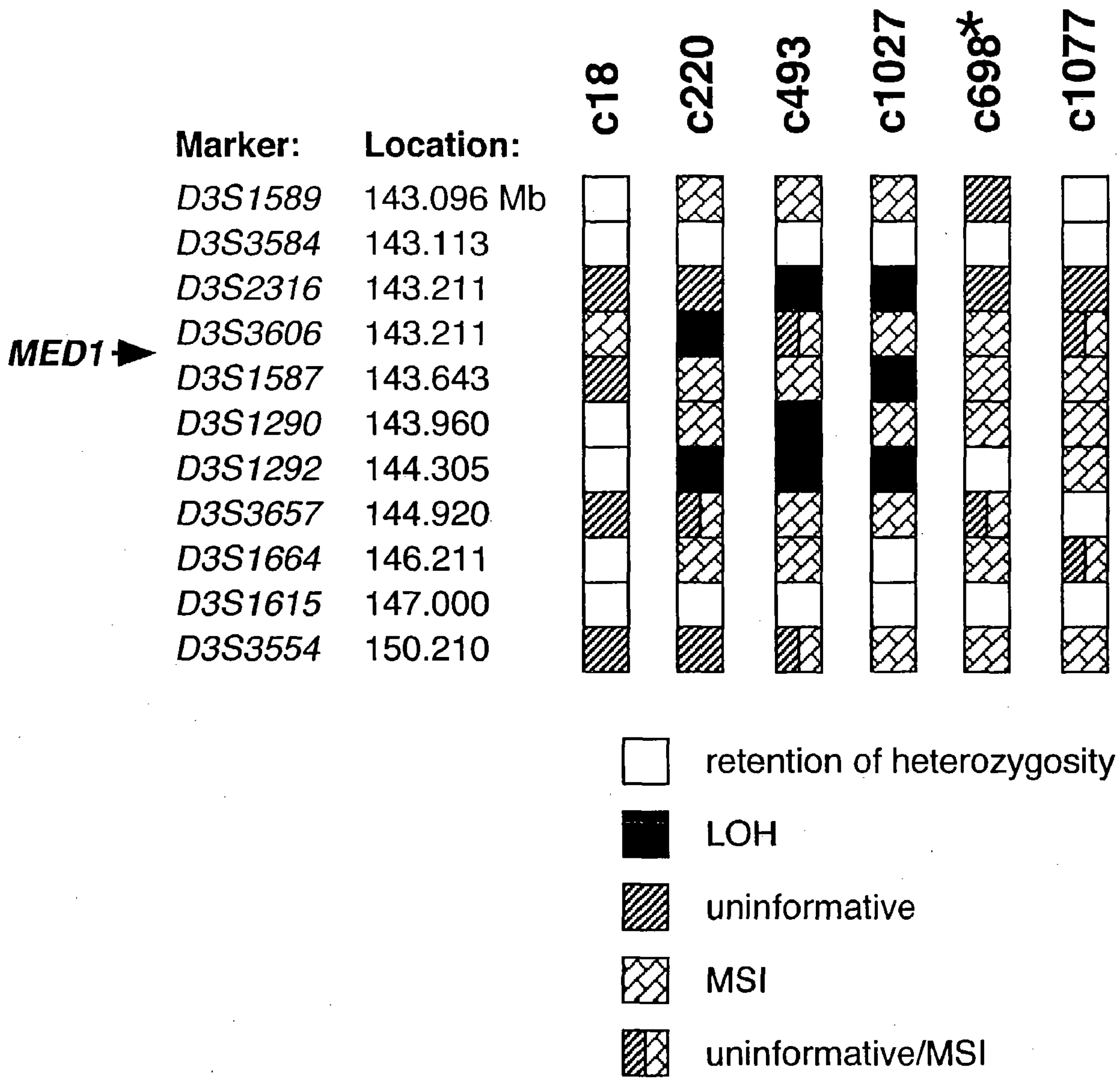

c18
c220
c493
c1027
c698*
c1077
Marker:
Location:
D3S1589
143.096 Mb
D3S3584
143.113
D3S2316
143.211
MED1
D3S3606
143.211
D3S1587
143.643
D3S1290
143.960
D3S1292
144.305
D3S3657
144.920
D3S1664
146.211
D3S1615
147.000
D3S3554
150.210
retention of heterozygosity
LOH
uninformative
MSI
uninformative/MSI

Fig. 20D

MED1 glycosylase assay

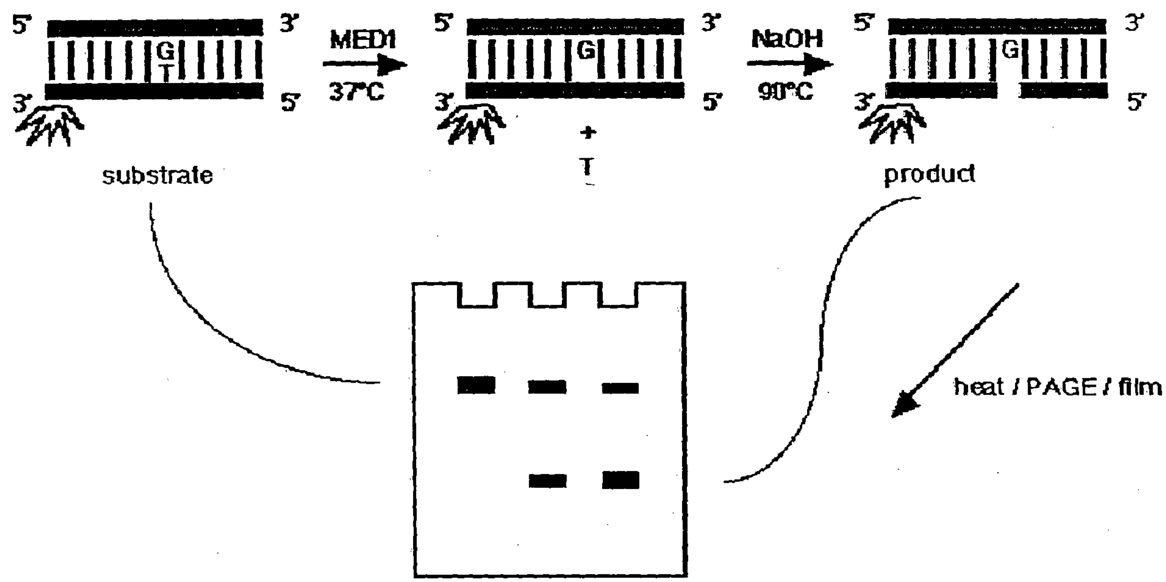
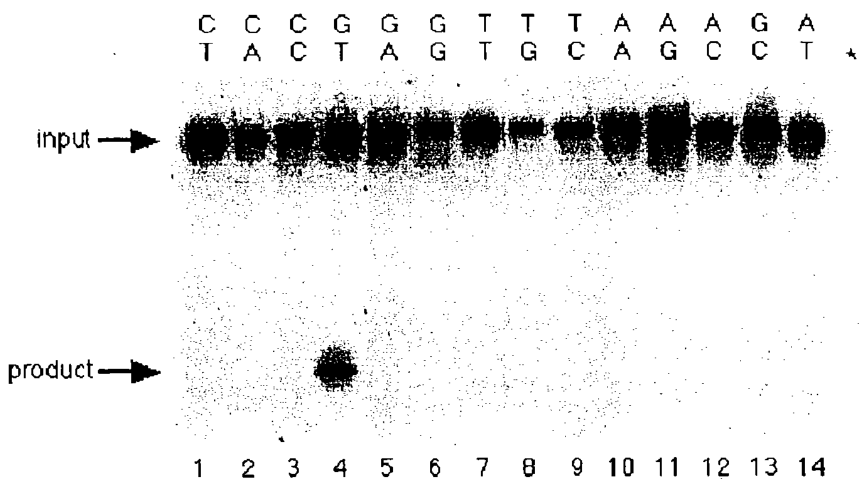


Fig. 21



* The asterisk indicates that the bottom oligonucleotide strand is radioactively labelled

Fig. 22

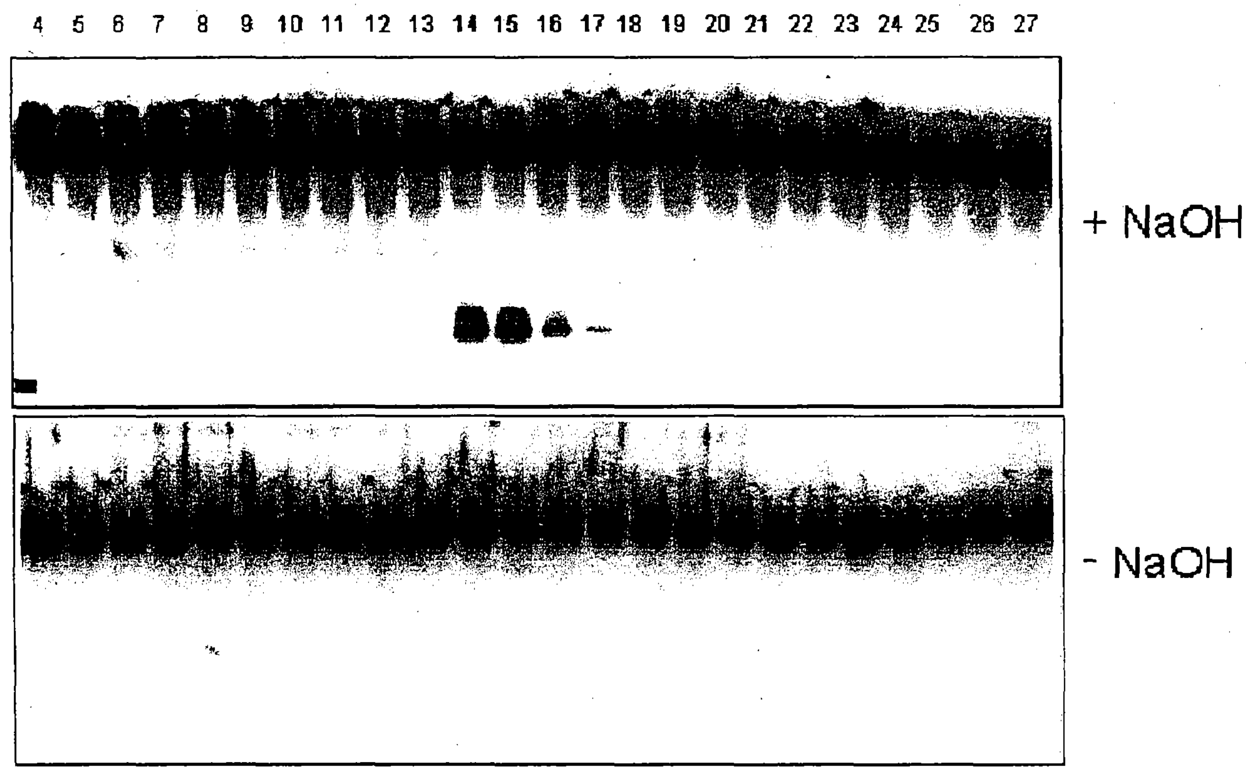

4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27
+ NaOH
- NaOH

Fig. 23

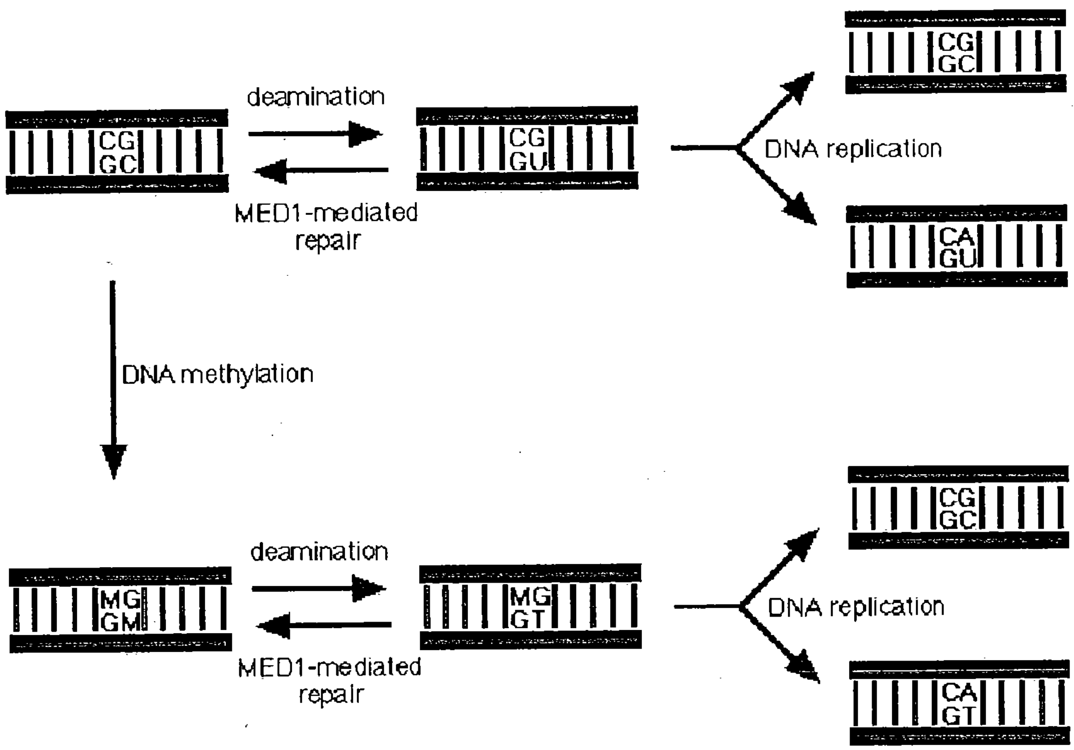

deamination
CG
GC
MED1-mediated
repair
CG
GU
DNA replication
CG
GC
CA
GU
DNA methylation
deamination
MG
GM
MED1-mediated
repair
MG
GT
DNA replication
CG
GC
CA
GT

Fig. 24

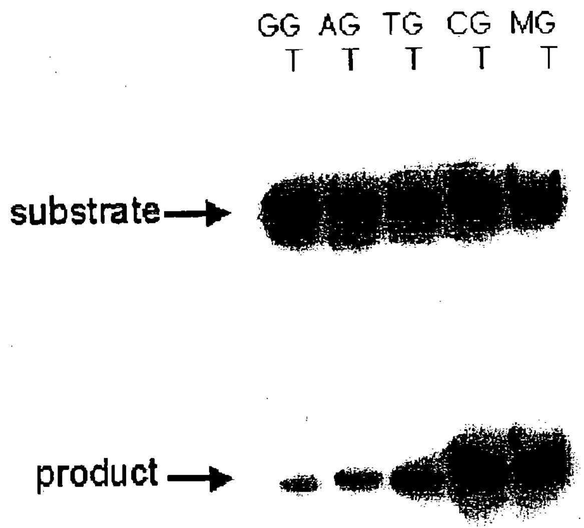


Fig. 25

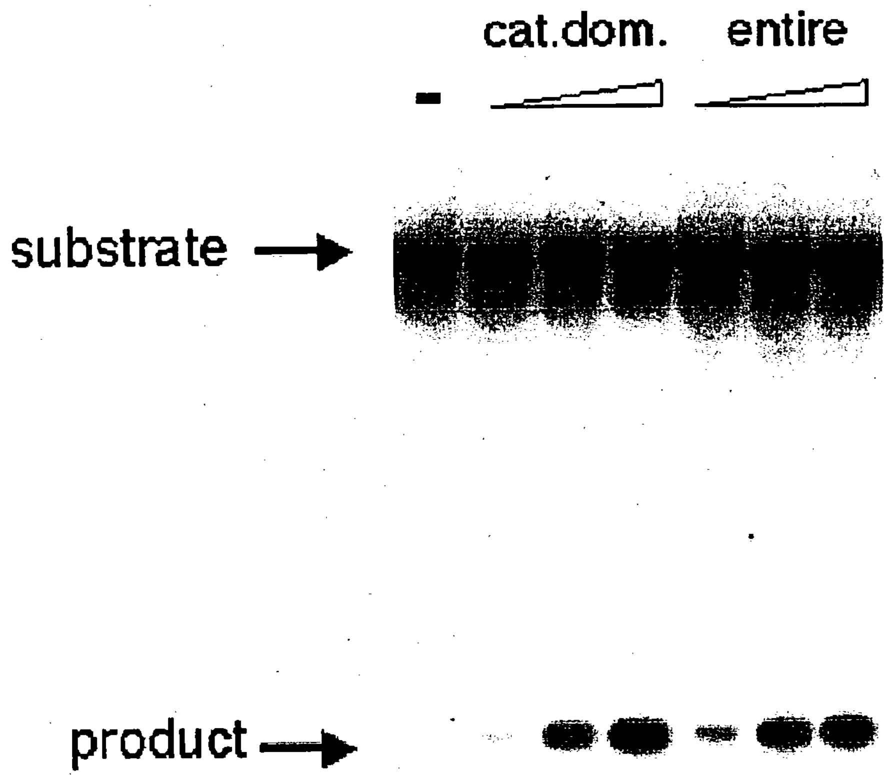
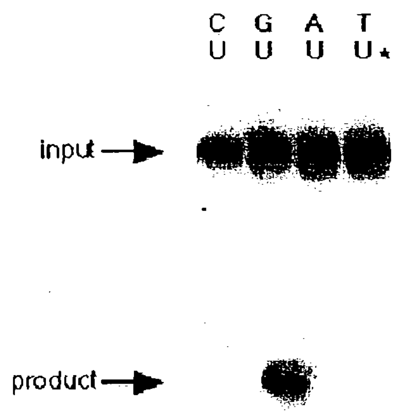


Fig. 26



*The asterisk indicates that the bottom oligonucleotide strand is radioactively labelled

Fig. 27

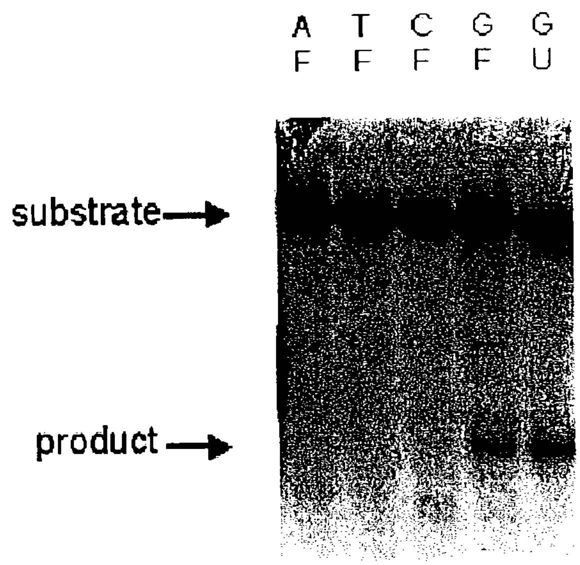
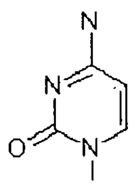
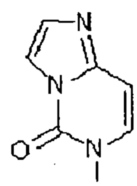


Fig. 28



Cytosine



3, N^4-Etheno-cytosine (E)

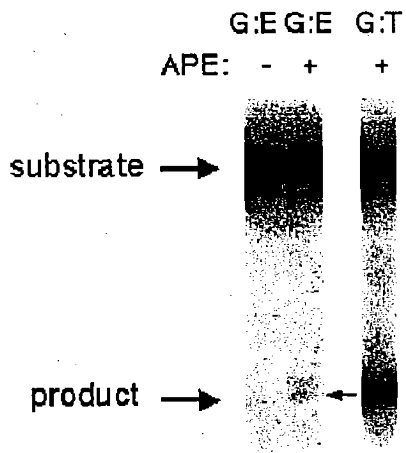


Fig. 29

1) Denature DNA fragments and mix with CT-SNP probe in order to generate heteroduplex

5′ ____CpG____ 3′ → 5′ ____CpG____ 3′ heteroduplex
3′ ____GpC____ 5′ → 3′ ____GpT____ 5′ with G:T mismatch

allele 1

or

5′ ____CpA____ 3′ → 5′ ____CpA____ 3′ homoduplex
3′ ____GpT____ 5′ → 3′ ____GpT____ 5′

allele 2

+

3′ ____GpT____ 5′

CT-SNP probe

2) Incubate annealed molecules with recombinant MED1 followed by NaOH in order to cleave heteroduplex

↓ + MED1 ↓
+ NaOH

5′ ____CpG____ 3′ or 5′ ____CpA____ 3′
3′ ____Gp ____ 5′ 3′ ____GpT____ 5′

cleaved heteroduplex uncleaved homoduplex

3) Separate fragments of the cleaved strand by standard techniques (e.g. by electrophoresis)

Fig. 30